# Defining the Strength of Evidence in Therapeutic Development

**CERSI Summit, Jan. 12, 2020**

**Steven Goodman, MD, PhD**

**Associate Dean, Professor of Epidemiology & Medicine**

**Perrine Janiaud, PhD**

**Post-doctoral fellow**

**Stanford University School of Medicine**

# Historical and regulatory context (1)

✓ 1938 Federal Food, Drug and Cosmetic Act ➔ drug product

  safe

✓ 1962 Kefauver-Harris amendment➔ drug product **safe and**

  **effect**  **Substantial evidence of effectiveness**: "evidence consisting of

  adequate and well controlled investigations,…"
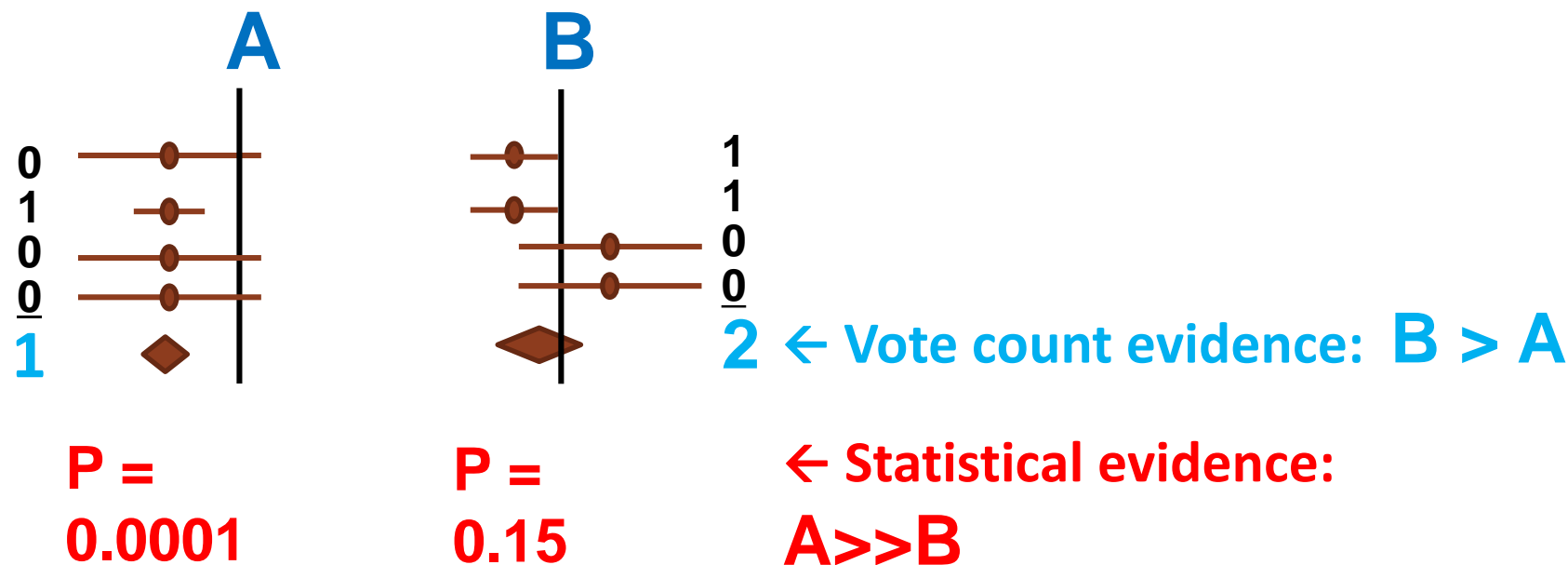
# Features of 2 "positive trials" rule

➢ **Aims to minimize probability of false positives due to bias or misleading evidence.**

➢ **Operationally clear**

**"…a single favorable study among several similar attempts that failed to support a finding of effectiveness would not constitute persuasive support for a product use unless there were a strong argument for discounting the outcomes in the studies that failed to show effectiveness."**

**FDA, 1998 Guidance on Clinical Evidence**

STANFORD
UNIVERSITY

# Features of 2 positive trials rule

➢ **Actually a form of meta-analysis (i.e. vote counting), albeit one that is statistically flawed.**



**A**

**B**

0
1
0
0
**1**

1
1
0
0
**2** ← **Vote count evidence:** **B > A**

**P = 0.0001**

**P = 0.15**

← **Statistical evidence:**
**A>>B**

STANFORD UNIVERSITY

# EFFECTS OF ADJUVANT TAMOXIFEN AND OF CYTOTOXIC THERAPY ON MORTALITY IN EARLY BREAST CANCER

## An Overview of 61 Randomized Trials among 28,896 Women

EARLY BREAST CANCER TRIALISTS' COLLABORATIVE GROUP

**Abstract** We sought information worldwide on mortality according to assigned treatment in all randomized trials that began before 1985 of adjuvant tamoxifen or cytotoxic therapy for early breast cancer (with or without regional lymph-node involvement). Coverage was reasonably complete for most countries. In 28 trials of tamoxifen nearly 4000 of 16,513 women had died, and in 40 chemotherapy trials slightly more than 4000 of 13,442 women had died. The 8106 deaths were approximately evenly distributed over years 1, 2, 3, 4, and 5+ of follow-up, with little useful information beyond year 5.
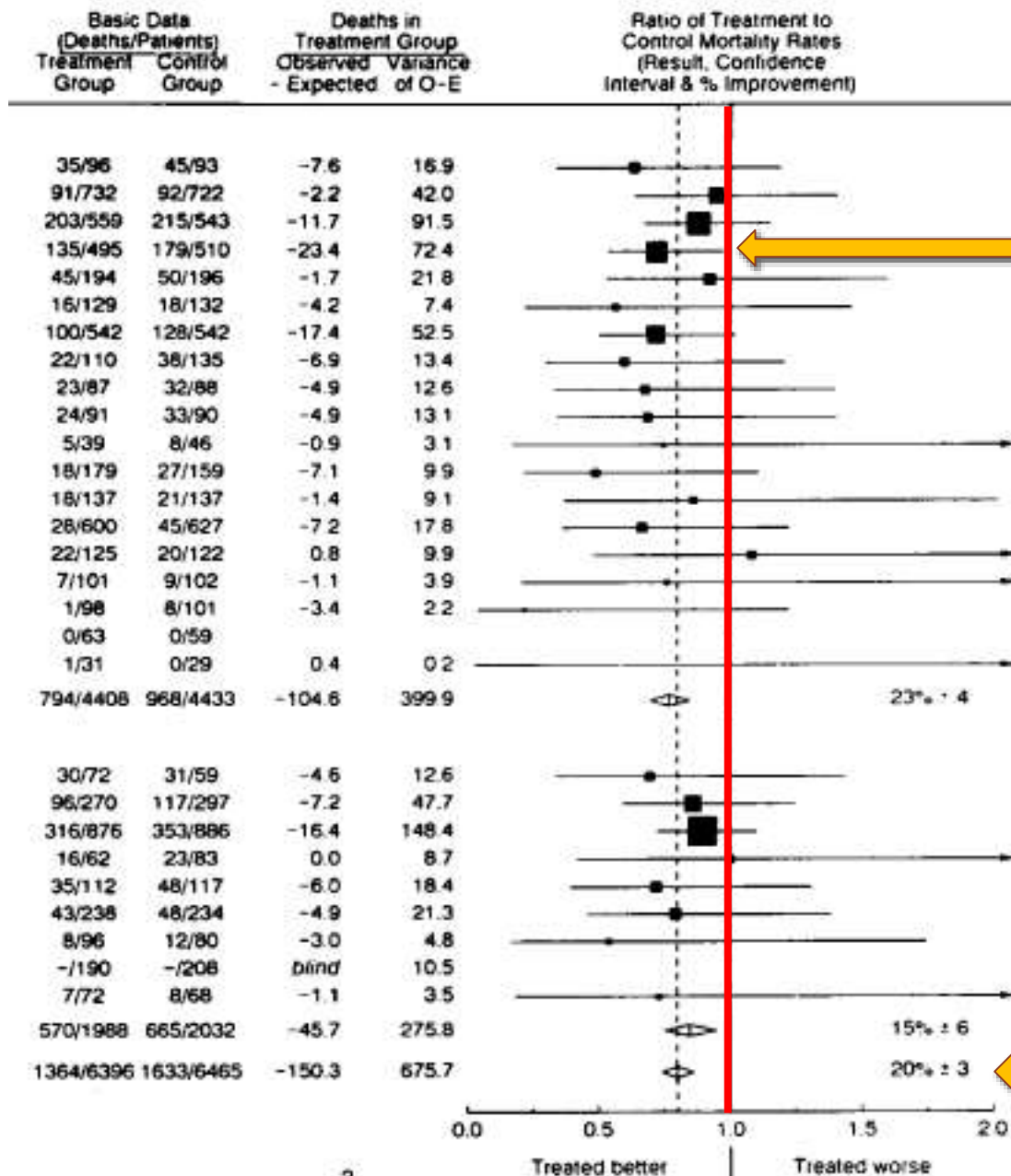
Systematic overviews of the results of these trials demonstrated reductions in mortality due to treatment that were significant when tamoxifen was compared with no tamoxifen ($P<0.0001$), any chemotherapy with no chemotherapy ($P = 0.003$), and polychemotherapy with single-agent chemotherapy ($P = 0.001$). In tamoxifen trials, there was a clear reduction in mortality only among women 50 or older, for whom assignment to tamoxifen reduced the annual odds of death during the first five years by about one fifth. In chemotherapy trials there was a clear reduction only among women under 50, for whom assignment to polychemotherapy reduced the annual odds of death during the first five years by about one quarter. Direct comparisons showed that combination chemotherapy was significantly more effective than single-agent therapy, but suggested that administration of chemotherapy for 8 to 24 months may offer no survival advantage over administration of the same chemotherapy for 4 to 6 months.

Because it involved several thousand women, this overview was able to demonstrate particularly clearly that both tamoxifen and cytotoxic therapy can reduce five-year mortality. (N Engl J Med 1988; 319:1681-92.)

**(B) Women aged ≥ 50 years at entry**

**Tamoxifen to reduce ER+ breast cancer mortality**

| Basic Data (Deaths/Patients) | | Deaths in Treatment Group | | Ratio of Treatment to Control Mortality Rates (Result, Confidence Interval & % Improvement) |
|---|---|---|---|---|
| Treatment Group | Control Group | Observed - Expected | Variance of O-E | |
| 35/96 | 45/93 | -7.6 | 16.9 | |
| 91/732 | 92/722 | -2.2 | 42.0 | |
| 203/559 | 215/543 | -11.7 | 91.5 | |
| 135/495 | 179/510 | -23.4 | 72.4 | |
| 45/194 | 50/196 | -1.7 | 21.8 | |
| 16/129 | 18/132 | -4.2 | 7.4 | |
| 100/542 | 128/542 | -17.4 | 52.5 | |
| 22/110 | 38/135 | -6.9 | 13.4 | |
| 23/87 | 32/88 | -4.9 | 12.6 | |
| 24/91 | 33/90 | -4.9 | 13.1 | |
| 5/39 | 8/46 | -0.9 | 3.1 | |
| 18/179 | 27/159 | -7.1 | 9.9 | |
| 18/137 | 21/137 | -1.4 | 9.1 | |
| 28/600 | 45/627 | -7.2 | 17.8 | |
| 22/125 | 20/122 | 0.8 | 9.9 | |
| 7/101 | 9/102 | -1.1 | 3.9 | |
| 1/98 | 8/101 | -3.4 | 2.2 | |
| 0/63 | 0/59 | | | |
| 1/31 | 0/29 | 0.4 | 0.2 | |
| 794/4408 | 968/4433 | -104.6 | 399.9 | 23% ± 4 |
| 30/72 | 31/59 | -4.6 | 12.6 | |
| 96/270 | 117/297 | -7.2 | 47.7 | |
| 316/876 | 353/886 | -16.4 | 148.4 | |
| 16/62 | 23/83 | 0.0 | 8.7 | |
| 35/112 | 48/117 | -6.0 | 18.4 | |
| 43/238 | 48/234 | -4.9 | 21.3 | |
| 8/96 | 12/80 | -3.0 | 4.8 | |
| -/190 | -/208 | blind | 10.5 | |
| 7/72 | 8/68 | -1.1 | 3.5 | |
| 570/1988 | 665/2032 | -45.7 | 275.8 | 15% ± 6 |
| 1364/6396 | 1633/6465 | -150.3 | 675.7 | 20% ± 3 |

0.0    0.5    1.0    1.5    2.0

Treated better | Treated worse

Test for heterogeneity: $X^2_{26} = 19.7$: NS

**Only 1 of 23 trials are significant**

**20% mortality reduction, $P < 10^{-9}$**

# Features of 2 positive trials rule

➤ **Imperfect evidence summary, leading to ad hoc adjustments w/o principled justification.**

➤ **Hard to establish clear or consistent FDA "case law".**

➤ **Incentivizes outcome switching, post-hoc chicanery, evidence suppression and gaming.**

STANFORD
UNIVERSITY

# Demonstrating Substantial Evidence of Effectiveness for Human Drug and Biological Products
## Guidance for Industry

***DRAFT GUIDANCE***

**This guidance document is being distributed for comment purposes only.**

Comments and suggestions regarding this draft document should be submitted within 60 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit electronic comments to https://www.regulations.gov. Submit written comments to the Dockets Management Staff (HFA-305), Food and Drug Administration, 5630 Fishers Lane, Rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft document, contact (CDER) Ei Thu Lwin, Office of New Drug Policy, 301-796-0728 or (CBER) Office of Communication, Outreach and Development, 800-835-4709 or 240-402-8010, ocod@fda.hhs.gov.

# Deviations from 2 trial requirement

➢ **Large multicenter trial**

➢ **Evidence from related indications**

➢ **Strong mechanistic support**

➢ **Established natural history**

➢ **Evidence from similar drug classes**

➢ **Serious outcomes with unmet medical need**

➢ **Rare diseases, disease subsets**

➢ **Efficacy trial is unethical**

STANFORD UNIVERSITY

# Judgments in ≤1 trial justification

- ➤ "**Persuasive**" statistical effect. [Renders 2nd trial "unethical".]

- ➤ "**Strongly positive**"

- ➤ "**Robust**" statistical effect

- ➤ "**Inconsistent**" statistical evidence

- ➤ "**Compelling**" efficacy results

- ➤ "**Strength of evidence**" (sans defn)

- ➤ "[efficacy] could be **fairly** and **responsibly** concluded by experts."

- ➤ "In all cases FDA must [conclude] that there is substantial evidence…however, the **degree of certainty** supporting such a conclusion may differ, depending on clinical circumstances."

- ➤ **N.B. No definition nor quantitative measure of "certainty"**

# Statistical considerations

"A typical criterion for concluding that a trial is positive (showed an effect) is a p value of < 0.05 (two sided). A lower p value, for example, would often be expected for reliance on a single trial. **For a serious disease with no available therapy or a rare disease … a somewhat higher p value – if prespecified and appropriately justified – might be acceptable.** "

*2019 Guidance, lines 602-8*

# Can a formal evidence score help?

➢ **Evidential strength modified by study design and drug features.**

- ❖ Large, multicenter RCT        Positive
- ❖ Observational        Negative
- ❖ Surrogate endpoints        Negative
- ❖ Short term outcomes        Negative
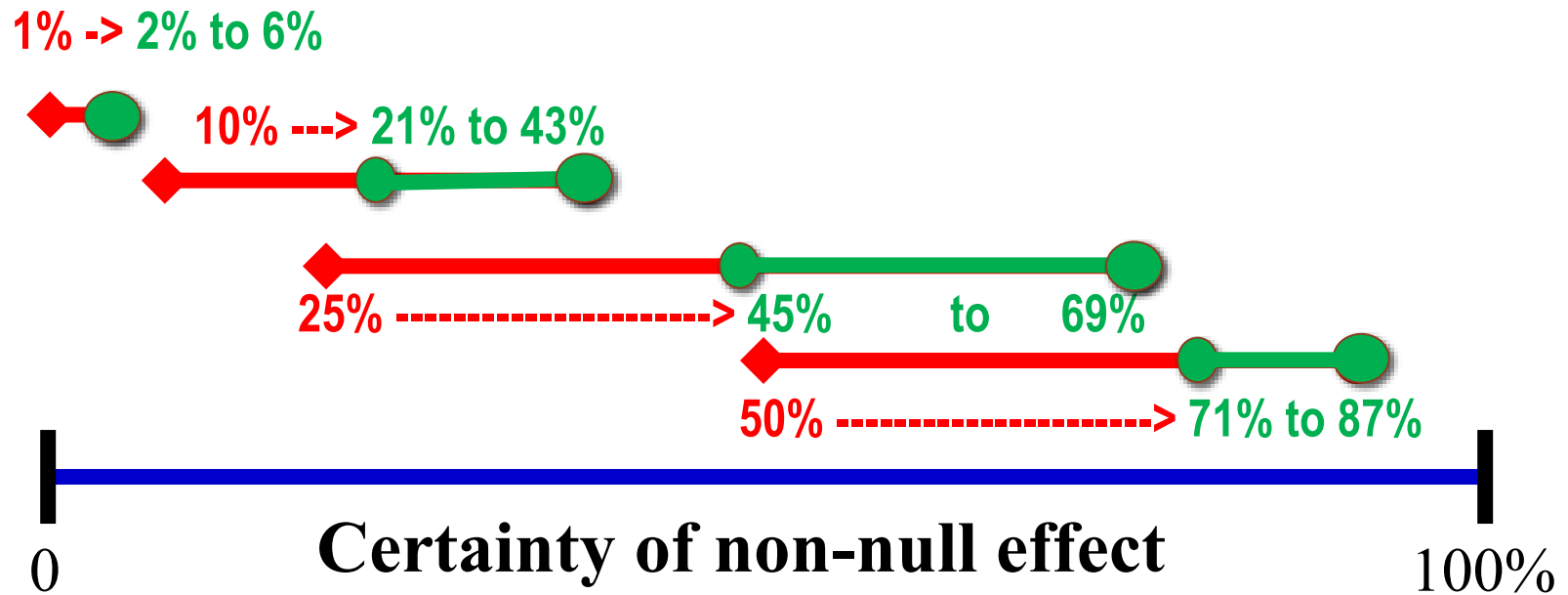- ❖ Validated mechanism        Positive

➢ **Certainty needed contingent on seriousness of disease and extant therapies.**

- ❖ Orphan disease w/o alternative:        Lower bar
- ❖ Me-too drug with multiple alternatives:        Higher bar.
- ❖ Breakthrough drug with better safety        Lower bar
- ❖ Serious condition        Lower bar
- ❖ Safety concern        Higher bar

# Creating an evidence score

➢ **Bayesian methods provide a valid calculus of evidence.**

➢ **An evidence score can incorporate the factors used to justify flexibility.**

➢ **Score is on quantitative scale that can be translated into:**

  ❖ "Certainty"

  ❖ P-values

  ❖ Prediction probabilities

  ❖ Bayes factors

# Translating P=0.05 into "certainty"



1% -> 2% to 6%

10% ---> 21% to 43%

25% -------------------> 45%     to     69%

50% -------------------> 71% to 87%

**Certainty of non-null effect**

0                                                                100%

# Source of cases: 2013-2018

All applications going through **multiple review cycles** for clinical efficacy issues.
- ✓ 17 different ACs in CDER
- ✓ 5 different ACs in CBER

**Documents:**
- FDA briefing documents and slides for the AC
- AC minutes
- AC transcripts (if needed)
- Approval letter and/or complete response letters
- Summary review used for the approval
- Office Director Memo

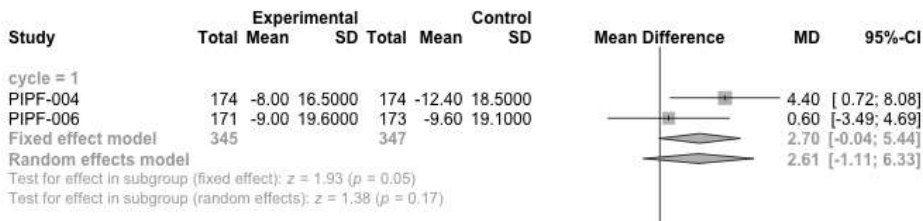| Characteristics | N |
|---|---|
| Reviewed by CDER and CBER | 912 |
| Receiving at least one CRL | 117 |
| **Clinical efficacy** | **22** |
| - Expedited | 11 |

STANFORD UNIVERSITY

# Pirfenidone (Ebriet) for idiopathic pulmonary fibrosis

STANFORD
UNIVERSITY

**Primary endpoint:** Mean change in % Predicted FVC (At 72 weeks)

| Cycle 1 | |
|---|---|
| **Evidence Score = 5.9** | |
| **1.9** | Statistical component |
| ✓ | Randomized |
| ✓ | Endpoint measured at 72 weeks (recommended by the FDA) |
| ⊘ | Surrogate endpoint |
| ✓ | Life-threatening disease / orphan |
| ⊘ | Mechanism of action not established |

| Study | Experimental Total Mean SD | Control Total Mean SD | Mean Difference | MD 95%-CI |
|---|---|---|---|---|
| cycle = 1 | | | | |
| PIPF-004 | 174 -8.00 16.5000 | 174 -12.40 18.5000 | | 4.40 [ 0.72; 8.08] |
| PIPF-006 | 171 -9.00 19.6000 | 173 -9.60 19.1000 | | 0.60 [-3.49; 4.69] |
| Fixed effect model | 345 | 347 | | 2.70 [-0.04; 5.44] |
| Random effects model | | | | 2.61 [-1.11; 6.33] |

Test for effect in subgroup (fixed effect): $z = 1.93$ ($p = 0.05$)
Test for effect in subgroup (random effects): $z = 1.38$ ($p = 0.17$)

### Cycle 1:
❖ Only one study statistically significant, unclear clinical relevancy
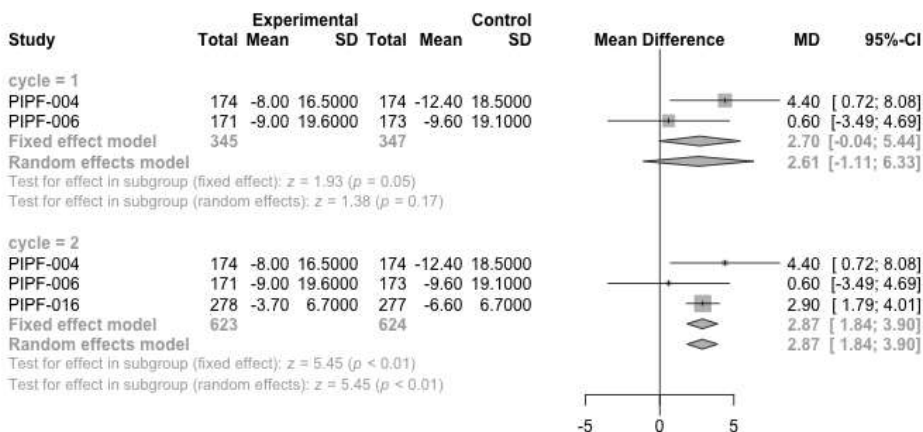❖ Life threatening disease ➔ mortality was a secondary outcome ➔ post-hoc analyses

**CRL**

STANFORD UNIVERSITY

**Pirfenidone CYCLE 2**
**Additional trial, P<0.01**

| | Experimental | | | Control | | | | |
|---|---|---|---|---|---|---|---|---|
| Study | Total | Mean | SD | Total | Mean | SD | MD | 95%-CI |
| **cycle = 1** | | | | | | | | |
| PIPF-004 | 174 | -8.00 | 16.5000 | 174 | -12.40 | 18.5000 | 4.40 | [ 0.72; 8.08] |
| PIPF-006 | 171 | -9.00 | 19.6000 | 173 | -9.60 | 19.1000 | 0.60 | [-3.49; 4.69] |
| Fixed effect model | 345 | | | 347 | | | 2.70 | [-0.04; 5.44] |
| Random effects model | | | | | | | 2.61 | [-1.11; 6.33] |
| Test for effect in subgroup (fixed effect): z = 1.93 (p = 0.05) | | | | | | | | |
| Test for effect in subgroup (random effects): z = 1.38 (p = 0.17) | | | | | | | | |
| **cycle = 2** | | | | | | | | |
| PIPF-004 | 174 | -8.00 | 16.5000 | 174 | -12.40 | 18.5000 | 4.40 | [ 0.72; 8.08] |
| PIPF-006 | 171 | -9.00 | 19.6000 | 173 | -9.60 | 19.1000 | 0.60 | [-3.49; 4.69] |
| PIPF-016 | 278 | -3.70 | 6.7000 | 277 | -6.60 | 6.7000 | 2.90 | [ 1.79; 4.01] |
| Fixed effect model | 623 | | | 624 | | | 2.87 | [ 1.84; 3.90] |
| Random effects model | | | | | | | 2.87 | [ 1.84; 3.90] |
| Test for effect in subgroup (fixed effect): z = 5.45 (p < 0.01) | | | | | | | | |
| Test for effect in subgroup (random effects): z = 5.45 (p < 0.01) | | | | | | | | |

Mean Difference: -5   0   5

| Cycle 2 | |
|---|---|
| **Evidence Score = 18.9** | |
| **14.9** | Statistical component |
| ✓ | Randomized |
| ✓ | Endpoint measured at 72 weeks (recommended by the FDA) |
| ⊘ | Surrogate endpoint |
| ✓ | Life-threatening disease / orphan |
| ⊘ | Mechanism of action not established |

**Director's memo:** "The clear efficacy of this drug and severity of IPF and lack of effective drugs establishes a clear risk-benefit assessment that allows approval."

APPROVED

STANFORD UNIVERSITY
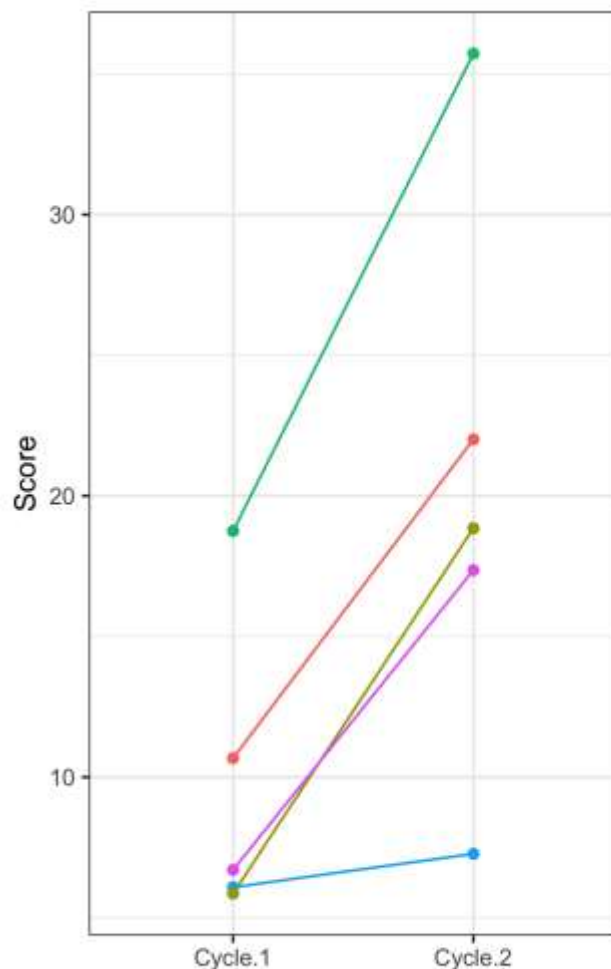
**Tolvaptan**: "In his statistical review, Dr. Lawrence notes that both the data and analysis quality "were excellent" in REPRISE and goes so far as to describe the trial as "an exemplar" for not only future trials in ADPKD, but for all clinical trials"

- *Summary review*

Applications

- ADDYI-flibanserin (Hypoactive sexual desire disorder) - Satisfactory sexual events
- ESBRIET-pirfenidone (Idopathic pulmocary fibrosis) - Predicted FVC
- JYNARQUE-tolvaptan (Autosomal dominant polycystic kidney disease) - Annualized change in eGFR (CKD-EPI)
- NORTHERA-droxidopa (Neurogenic orthostatic hypotension) - Dizziness
- XIIDRA-lifitegrast (Dry eyes disease) - Mean change in Eye Dryness Score

**Droxidopa**: "There is no doubt that the data are at the "margin" for approvability"

- *Office director*

# Conclusion

➢ **Evidence score starts with a formal meta-analysis of extant evidence, a valid method of evidence combination. May require more complete submission packages if "negative" studies count.**

➢ **The qualitative factors in the evidence score are all already used, albeit informally.**

➢ ***Evidence score does not replace judgment*, but improves rigor, transparency and consistency of FDA efficacy assessments.**

# Thanks to:

- **<u>FDA Collaborators</u>: Estelle Russek-Cohen, Telba Irony, [Tom Permutt and Lisa LaVange]**

- **Janet Woodcock**

- **CERSI**

- **METRICS (Meta-research Innovation Center @ Stanford)**

STANFORD UNIVERSITY